

Auditing Lymph Node Detection Failures in a Multimodal 3D nnU-Net

A Radiomic Analysis of Biological Boundaries and
Algorithmic Triggers in Head and Neck Cancer

Master's Thesis

Bioinformatics

Authors

Anders Lydig Kristensen (Student ID: 201808522)

Esther Helga Klemenzardóttir (Student ID: 202007451)

Advisors

Juraj Bergman, *Tenure Track Assistant Professor, Aarhus University*

Stine Sofia Korreman, *Professor of Medical Physics, Aarhus University*

Hospital

Department of Molecular Biology and Genetics

Bioinformatics Research Center

Aarhus University

June 2026

Abstract

In head and neck squamous cell carcinoma (HNSCC), multimodal deep learning models are currently being investigated to assist in the delineation of target volumes. A 3D nnU-Net, evaluated within the PROSA-GTV clinical trial, showed high accuracy for macroscopic primary tumors but exhibited a systematic false-negative rate for regional lymph nodes. This thesis utilizes standardized radiomics as a mathematical proxy to audit the algorithm’s latent decision-making, separating its learned triggers from physical tissue characteristics. Utilizing clinical data from the trial (CT, PET, T1, and T2 MRI), Generalized Estimating Equations (GEE) were applied to isolate the confounding variable of target volume. Subsequently, a Dual-Target Framework employing Random Forest classifiers and out-of-fold permutation feature importance mapped the Algorithmic Boundary (active network segmentations) against a limited Biological Boundary (true malignancies versus a proxy of network-generated false positives). The analysis indicates that the nodal detection failure is driven by spatial voxel imbalances and compounded by deterministic post-processing (2D Z-axis erosion filters), rather than an inherent inability to process nodal biology. Furthermore, the Random Forest models reveal that the network relies on macroscopic metabolic shortcuts, specifically ^{18}F -FDG PET signals, to trigger segmentations. Consequently, the network likely underutilizes discriminative spatial textures—such as continuous CT low-density matrices—that demonstrate physical separation between true malignancies and background computational artifacts. The data suggests the 3D U-Net exhibits misaligned data priorities due to shared convolutional filters and a volumetric target imbalance. Resolving these blind spots could require decoupling the primary tumor and nodal architectures, transitioning to late-fusion to mitigate metabolic overshadowing, and implementing minority-class patch oversampling or dynamically weighted loss functions to mathematically prioritize micro-volumes.

Acknowledgements

We thank Juraj Bergman for his assistance with the statistical design, methodological decision-making, and the structural organization of this thesis. We thank Stine Sofia Korreman for providing the clinical dataset, supplementary materials, and necessary clinical context. We also thank both for their constructive feedback and discussions throughout the project.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 2 | Clinical Context and the Deployed 3D U-Net | 2 |
| 3 | Literature Review | 3 |
| 3.1 | Convolutional Networks and Radiomic Feature Extraction | 3 |
| 3.2 | Biological and Radiographic Distinction Between Targets | 4 |
| 3.3 | Modality Imbalance and Size-Dependent Detection Bias | 4 |
| 3.4 | Radiomics as a Quantitative Proxy for Network Behavior | 5 |
| 3.5 | Methodological Limitations of Radiomic Proxies | 5 |
| 4 | Problem Statement | 6 |
| 5 | Methods | 7 |
| 5.1 | Data Preparation, Quality Control, and Radiomic Extraction | 7 |
| 5.2 | Volumetric Bias Analysis | 10 |
| 5.3 | Feature Sanitization and Confounder Isolation | 12 |
| 5.4 | Unsupervised Clustering and Medoid Selection | 13 |
| 5.5 | The Dual-Target Framework: Modeling Biological and Algorithmic Boundaries | 13 |
| 5.6 | Software and Computational Environment | 17 |
| 6 | Results | 18 |
| 6.1 | Cohort Yield and Target Distributions | 18 |
| 6.2 | Volumetric Bias: Target Size and Adjusted Detection Performance | 19 |
| 6.3 | Feature Sanitization Yield and Dimensionality Reduction | 22 |
| 6.4 | Global Performance of Algorithmic and Biological Boundaries | 25 |
| 6.5 | Feature Importance: Biological and Algorithmic Boundaries | 27 |
| 7 | Discussion | 33 |
| 7.1 | Volumetric Bias and Anatomical Voxel Imbalance | 33 |
| 7.2 | The Algorithmic Boundary: Over-Reliance on Metabolic Confounders | 34 |
| 7.3 | The Biological Boundary: Underutilized Structural Features and Phenotypes | 36 |
| 7.4 | Random Forest Ablation Mechanics | 38 |
| 7.5 | Recommendations for Model Optimization | 39 |
| 7.6 | Methodological Limitations and Future Work | 41 |
| 8 | Conclusion | 45 |